



论文发表了,数据去哪儿了?

黄勇平代表:避免科研“隐形浪费”

□ 记者 陈颖婷

两会 关注

“一旦国外关闭这样的数据集,我们的科研将会受到极大的掣肘。”全国人大代表、上海交通大学环境科学与工程学院特聘教授黄勇平在今年全国两会上带来了一份关于加强高质量数据集建设的建议。他指出,目前我国科研领域广泛使用的数据集仍高度依赖国外数据库——生物医药领域的NCBI、Pubmed等国际平台几乎是科研人员每天都要访问的“地方”,这种局面隐藏着巨大的风险。

“一方面,我们把自己的大量数据上传到国外数据库,将宝贵的资源对外公开;另一方面,如果国外数据库关闭,我们的科研将受到极大掣肘。”黄勇平说,国际竞争日益激烈,建设中国自己的高质量数据集刻不容缓。

“重硬轻软”的困局:有高水平设备投入,但运维人员不足

“数据集建设包括收集、储存以及挖掘应用,否则就不会发挥作用。只有让‘数据流动起来’,高质量数据集建设才能真正落地生根。”黄勇平说。调研中,黄勇平发现,当前我国科学数据建设面临三大痛点。

首先是科学数据分散、标准不统一。传统科研过程中,数据多由各个小团队私有管理、分散存储,主要通过论文形式发布数据分析利用的成果,难以实现高质量数据的高水平复用。

其次是投入比例严重失衡。主管部门也会投入来支持数据集建设,但“科研项目资金投入存在‘重硬轻软’现



全国人大代表黄勇平

张弛 摄

象。”黄勇平直言,由于科学数据的价值不像硬件设备那样直观可见,传统项目资金主要投向数据设备,对数据管理的人力投入比例极低,导致“有高水平设备,却没有足够人力来运营维护”。一些

前期投入的项目甚至因配套的数据治理和存储资金不足,大量科学数据来不及开发即被删除,造成科研资源严重浪费。

第三是数据集建设缺乏激励机制。当前科研评价仍以“论文数量、影响因子、引用次数”为核心,数据仅被视为论文的“附属品”。“高质量科学数据集无法作为高水平科研成果得到学界认可,科研人员自然没有动力持续投入开发。”黄勇平表示,科学数据作为知识载体的复用价值得不到认可,只能通过论文进行一次价值发掘,无法满足数据驱动的大规模科研协作需要。

代表开“药方”:让数据集建设者的工作同样能够得到认同

针对上述问题,黄勇平提出全面规

划布局高质量科学数据集。国家要从数据安全角度切实重视高质量数据集建设,以应用需求为导向,针对重点学科领域,依托国家实验室体系和重点企业实验室,分类分级做好规划布局。

同时,加大科学数据项目资金投入。黄勇平建议设立专项基金,聚焦重点学科领域启动一批重大科研项目。科研项目预算应全面设置“科学数据处理”专属科目、单独列支,提高科学数据治理投入比例。

“要将科学数据贡献纳入学术评价体系,设置‘数据贡献度’指标。”黄勇平建议,将“科学数据出版、数据复用次数、数据衍生成果”纳入职称评审、项目结题、人才申报的核心指标,使科学数据贡献者与论文作者享有同等学术认可。



建设资金要管好 优选建行沪资宝

70余年造价根基 铸就数字化资金监管标杆

